Global Standards for Big Geospatial Data, the Copernicus challenge

**Peter Strobl, Michael Lutz, Jordi Escriu, Matt Purss**

INSPIRE 2017 conference
Strasbourg, France

_Peter.Strobl@ec.europa.eu_
Copernicus

• The EU’s *Earth Observation* program (one of the biggest globally)

• A major sponsor of *EO related services*

• An incubator for the *digital economy*

• *its main results are large amounts of geospatial data* of which the vast majority consists of either remotely sensed imagery or its derivatives.

... (almost) all *free and open!*
Challenge? – what challenge?!


'Copernicus data should be compliant with Member States' spatial reference data as well as with implementing rules and technical guidelines of the infrastructure for spatial information in the Union established by Directive 2007/2/EC of the European Parliament and of the Council’
Copernicus & INSPIRE

As presented earlier:
- Metadata
- Cataloguing
- Web services
- ...

are (largely) INSPIRE compliant

... so, what else?

<table>
<thead>
<tr>
<th>Requirement</th>
<th>Discovery Service</th>
<th>View Service</th>
<th>Download Service</th>
</tr>
</thead>
<tbody>
<tr>
<td>OGC Catalogue Service Web 2.0.2</td>
<td>Yes</td>
<td>Yes</td>
<td></td>
</tr>
<tr>
<td>ISO 19115 Geographic Information - Metadata</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>ISO 19119 Geographic Information - Services</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>ISO 19139 Geographic Information - Implementation specification</td>
<td>Yes</td>
<td></td>
<td></td>
</tr>
<tr>
<td>OGC Web Map Service 2.0.0</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Specific INSPIRE requirements</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

INSPIRE 2016 conference
30 September 2016
Earth observation data are geophysical measurements:

- of a well defined spatial area
- using parts of the electromagnetic spectrum,
- taken at regular intervals in space, and
- with a certain repeat cycle

In all these dimensions discretisation and quantisation is a prerequisite to deal with these data in a digital way.

... in case of a regular and orthogonal spatial order they are commonly often referred to as ‘images’ or ‘raster data’
INSPIRE & Raster Data, Grids, Coverages

INSPIRE addresses raster data in the following chapters:

- Annex I Reference Grids
- Annex II: Elevation, Orthoimagery, Landcover

Provided solutions include a selection of the following:

- Grid_ETRS89-LAEA, 5 levels: $10^n \text{m}$, $n=0,1$, i.e. 1m-100 km
- Grid_ETRS89-GRS80zn, ’zoned geographic’ grid, 25 levels: 0.1m-120km
- any grid based on ETRS89-LCC, ETRS89-TMzn

... and Copernicus?
Copernicus ‘real-world’ data grids

Georeferenced data in the portfolio:

- EL: EU-DEM: 25m ETRS89-LAEA
- OI: Sentinel-2: 10m-20m-60m, WGS84-UTM
- Sentinel-3: 300m, 1km, Plate-Caree
- LC: CLMS HighResolutionLayers: 20m ETRS89-LAEA

Standards: Maybe – somehow ...
Harmonisation: Certainly not!

Data rates growing exponentially:
Now ~6 soon >9TB/day only Copernicus!
Why are different grids a problem?

Because conversion between grids:
- always entails an interpolation of data
- always diminishes data accuracy
- is always irreversible
- it’s effects are cumulative
- depending on algorithm could become very computer-intense
… it would be highly desirable that all the themes with similar needs make use of the same geographical grid system in order to maintain their coherence.”
Who is to solve the problem of gridded data?

- Providers (NASA, ESA, EUMETSAT, INPE, ...)
- Analysts and Services (Administration, Research, Industry)
- Users (different thematic fields)
- Standards Authorities (ISO, OGC, EC-INSPIRE, ...)

Let’s have a workshop!
Where are the grids?

Sensor A
- Preprocessing
- Analysis
- Visualisation
- User 1

Sensor B
- Preprocessing
- Analysis
- Visualisation
- User 2

Sensor C
- Preprocessing
- Analysis
- Visualisation
- User 3
Which grids to address?

- Acquisition grids can’t be harmonised as they are volatile
- Display formats can’t be harmonised as they need to serve different purposes
- However, grids in data analysis must be harmonised to facilitate interoperability, preserve quality, and optimise efficiency
Which are the base requirements for a global analytical grid system?

- Earth fixed or plate fixed (which)
- Dimensionality, $3^{rd}$ (height), $4^{th}$ (time)
- Coverage (gaps)
- Uniqueness (overlaps)
- Equal area cells
- Discrete Cell hierarchy
- Cell identifiers
- Suitability for HPC (parallelisation)
- Suitability for the Cloud and for virtualisation
- Openness
- ...
What are the major external drivers?

- Copernicus is establishing the first consistent free and open global ortho-reference at 10m scale
  - Major reprocessing will occur!
- Currently no consistent (seamless and complete) grid exists to adequately hold these data.
- The EC is about to finance 4-5 “Data Information and Access Services” (DIAS) for Copernicus
  - Heterogeneity in dissemination will increase!
- Currently no common grid system is enabling the interoperability between them.
How do we proceed?

- Get all stakeholders on board
- Thoroughly analyse possible solutions (e.g. DGGS)
- Implement pilot studies and test beds (e.g. OGC)
- Facilitate transition and uptake

The basis of each data cube is a grid, the basis of a global data cube needs to be a global grid!

Thank you!